AD-A034 114    DRC INVENTORY RESEARCH OFFICE PHILADELPHIA PA       F/G 12/1
             HEURISTIC INTERPRETATION OF ANALYSIS OF COVARIANCE.(U)
             DEC 76   E GOTWALS

UNCLASSIFIED          IRO-TN-77-1                          NL

1 OF 1
AD-A
034 114

END
DATE
FILMED
2-15-77
NTIS

HEURISTIC INTERPRETATION OF ANALYSIS
OF COVARIANCE

DRC INVENTORY RESEARCH OFFICE
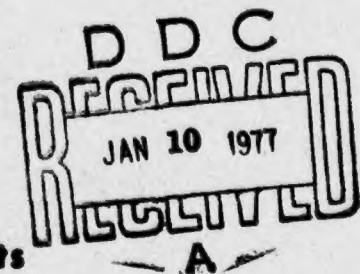PHILADELPHIA, PENNSYLVANIA

DECEMBER 1976

011175

# HEURISTIC INTERPRETATION OF ANALYSIS OF COVARIANCE

US ARMY
DARCOM

D R C
I N V E N T O R Y
R E S E A R C H
O F F I C E

December 1976

D D C

RECEIVED
JAN 10 1977

A

ROOM 800

U.S. CUSTOM HOUSE

2nd and Chestnut Streets

Philadelphia Pa. 19106

Information and data contained in this document are based on input available at the time of preparation. Because the results may be subject to change, this document should not be construed to represent the official position of the U.S. Army Materiel Command unless so stated.

i

A HEURISTIC INTERPRETATION OF THE ANALYSIS OF COVARIANCE


TECHNICAL REPORT

BY

EDWIN GOTWALS,III

DECEMBER 1976

DRC INVENTORY RESEARCH OFFICE
US ARMY LOGISTICS MANAGEMENT CENTER
US CUSTOM HOUSE          ROOM 800
2ND & CHESTNUT STREETS
PHILADELPHIA, PA. 19106

ii

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE *(and Subtitle)* <br><br> A HEURISTIC INTERPRETATION OF THE ANALYSIS OF COVARIANCE | | 5. TYPE OF REPORT & PERIOD COVERED <br><br> Technical Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) <br><br> Edwin Gotwals, III | | 8. CONTRACT OR GRANT NUMBER(s) |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS <br> DRC Inventory Research Office, ALMC <br> US Custom House, Room 800 <br> 2nd & Chestnut Sts., Phila., Pa. 19106 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS <br> US Army Materiel Development & Readiness Command <br> 5001 Eisenhower Avenue <br> Alexandria, VA  22333 | | 12. REPORT DATE <br> December 1976 |
| | | 13. NUMBER OF PAGES <br> 2b |
| 14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* | | 15. SECURITY CLASS. *(of this report)* <br><br> UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*


Approved for Public Release; Distribution Unlimited


17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*




18. SUPPLEMENTARY NOTES
      Information and data contained in this document are based on input available at the time of preparation.  Because the results may be subject to change, this document should not be construed to represent the official position of the US Army Materiel Development & Readiness Command unless so stated.

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

      Analysis of Covariance
      Regression Analysis
      Analysis of Variance
      Model Selection

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*
      A graphical approach is presented to describe the Analysis of Covariance to the analyst with a moderate statistical background.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

## TABLE OF CONTENTS

# A HEURISTIC INTERPRETATION OF THE ANALYSIS OF COVARIANCE

## 1. Introduction

The purpose of this paper is to provide an explanation of the Analysis of Covariance to the analyst who has a moderate statistical background, and who particularly is familiar with the terminology of Analysis of Variance. In so doing, graphical techniques for analyzing data are described which assist the analyst in selecting the underlying model which describes the relationship of the response variable to the predictor or treatment variables. It is felt that graphical techniques of this type provide much insight which is normally lost in traditional analysis. The Analysis of Covariance model is discussed as a general model form of the three familiar models - simple mean, linear regression, and Analysis of Variance. This description is intended to clarify the "cook book" or "black box" description of the Analysis of Covariance common in many statistical texts.

## 2. Motivation

Suppose in the course of a simple one way analysis of variance experiment designed to analyze the effects of p treatments on a response variable Y, a nuisance variable X is observed which is correlated with Y and varies between the runs of the experiment. The variability of X between each run confounds the traditional analysis of variance results, i.e., the potential resulting treatment effect may be attributed to either the difference in the treatment levels or to the variation of X between the treatments or to both.

Example: For a recent project at IRO we wanted to determine the effects of usage (miles driven in a given period) and odometer (the accumulated mileage at the beginning of the period) on the demand rate for replacement parts during the initial two-year history for the 2-1/2 ton truck. Since the data was collected before any experimental design could be considered, the various blocking schemes designed to control the investigative variables were not imposed on the data collection. Our initial results, looking at each variable independently indicated a

2

potential linear relationship with odometer and possibly a treatment
effect with usage. Upon further investigation, it was found that
there were different odometer levels associated with each treatment
level (one way analysis of variance). This became obvious when a two way
analysis of variance was tried and we found that odometer could not be con-
trolled within each treatment, i.e., we did not have observations for the
low usage and high odometer cells and similarly for the high usage and
low odometer cells. To determine if the usage effect found in the one way
analysis of variance was due to usage, or odometer, or both, an analysis of
covariance was tried where the odometer was considered the nuisance or un-
controllable variable and usage was the treatment variable. The results
from this analysis can be found in [2].

3. <u>Description of General Model</u>

Analysis of covariance uses regression analysis to remove the effect
of the nuisance variable from the response before applying the analysis of
variance. By so doing, only the true treatment effects are tested. The
underlying model is:

$$Y_{ij} = \mu_i + \beta(X_{ij} - \overline{X}..) + \epsilon_{ij} \tag{1}$$

$$i = 1,2,\ldots p \text{ treatments}$$

$$j = 1,2,\ldots n_i \text{ replications}$$

$$N = \sum_{i=1}^{p} n_i = \text{total sample size}$$

where $\mu_i$ is the true $i^{th}$ treatment mean

and $\beta(X_{ij} - \overline{X}..)$ is the linear relationship with the

covariate X

and $\epsilon_{ij}$ is the random noise term which is normally distributed

about the mean 0 with variance $\sigma^2$.

If this model is felt to appropriately describe the data – the next
section deals with ways to determine this – then Analysis of Covariance
merely provides a specific technique for estimating $\beta$ so that the response

3

variable Y may be corrected for the effect of X.

## 4. Model Selection

This section describes a graphical method for analyzing data in order to select an appropriate model. We are selecting from four possible models of which (1) is the general form and the other three specific cases. The intent of the graphical analysis is to offer the analyst an alternative way to look at his data which is more intuitive and discerning than conventional Analysis of Covariance methods. Basically, the technique involves a sequential review of various scatter diagrams for specific patterns implied by each of the models. The approach here is to consider the following models as degenerate cases of (1).

$$Y_{ij} = \mu + \epsilon_{ij} \qquad \begin{aligned} &i = 1,2,\ldots p \\ &j = 1,2,\ldots n_i \\ &N = \sum_i n_i \end{aligned} \qquad (2)$$

(mean model)

$$Y_{ij} = \alpha + \beta X_{ij} + \epsilon_{ij} \qquad \begin{aligned} &i = 1,2,\ldots p \\ &j = 1,2,\ldots n_i \\ &N = \sum_i n_i \\ &\text{and } \alpha = \mu - \beta\overline{X}.. \end{aligned} \qquad (3)$$

(regression model)

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad \begin{aligned} &i = 1,2,\ldots p \\ &j = 1,2,\ldots n_i \\ &N = \sum_i n_i \end{aligned} \qquad (4)$$

(analysis of variance model)

Each of the models are equivalent to (1) under special conditions, i.e.,

(1) = (2)  iff $\mu_i = \mu_\ell$ for every $i, \ell$ and $\beta = o$

(1) = (3)  iff $\mu_i = \mu_\ell$ for every $i, \ell$ and $\alpha = \mu - \beta\overline{X}..$

(1) = (4)  iff $\beta = o$

4

By making statistical estimates of each of the parameters in each
model, the prediction errors may be calculated. Assuming that a model ade-
quately describes the relationship, these errors should be random with mean
zero and the mean of their squares (MSE) minimum over all other models. Since
the models (2), (3), and (4) are degenerate cases of (1), the adequacy of these
models can be determined by comparing their MSEs and graphs of the errors
with the similar statistics using model (1).

The underlying difference between the models is captured in the
procedures for estimating the parameters. For the degenerate models the
$\mu_i$ and $\beta$ are estimated independently of each other, whereas in model (1)
these parameters are simultaneously estimated.

The goal in model selection is to determine the simplest model which
adequately describes the data. With this intent, the next four sections
considers each model as a null hypothesis in order of complexity. Under the
"Estimation" subheading, the null hypothesis is assumed correct and estimates
of the model parameters are derived along with calculations for the pre-
diction errors, and MSE. The next subheadings contain more complicated
alternative models in which additional variables are incorporated into the
null hypothesis model. In each of these cases, graphical techniques are
described by which the analyst can determine if he is making a type II
error by accepting the null hypothesis when the alternative model should
have been used. The "Conclusions" gives an overview of the null hypothesis
model.

5. <u>Null Hypothesis:Model (2)</u>   $(Y_{ij} = \mu + \varepsilon_{ij})$

Estimation

    (5.1) estimate $\mu$ by $\overline{Y} = \sum\limits_{i} \sum\limits_{j} Y_{ij}/N$

    (5.2) compute $_2\varepsilon_{ij} = Y_{ij} - \overline{Y}$ as the prediction error assuming
                                     model 2

    (5.3) compute $MSE_2 = \sum\limits_{i} \sum\limits_{j} (Y_{ij} - \overline{Y})^2/N-1$

5

<u>Suppose Model (3) is Correct</u> (but Model (2) is used)

(5.4)   then by replacing $Y_{ij}$ with (3) yields

$$_2\epsilon_{ij} = \mu - \beta\overline{X}.. + \beta X_{ij} + \epsilon_{ij} - \overline{Y}$$

but $\overline{Y} = \mu - \beta\overline{X}.. + \beta\overline{X}.. + \overline{\epsilon}..$

hence $_2\epsilon_{ij} = \beta(X_{ij} - \overline{X}..) + \epsilon_{ij} + \overline{\epsilon}..$

and $E(_2\epsilon_{ij}) = \beta(X_{ij} - \overline{X}..)$

(5.5) from (5.4) a plot of $_2\epsilon_{ij}$ against $X_{ij}$ would yield random
errors about the line $Y = \beta(X_{ij} - \overline{X}..)$

<u>Suppose Model (4) is Correct</u> (but Model (2) is used)

(5.6)   then by replacing $Y_{ij}$ with (4) yields

$$_2\epsilon_{ij} = Y_{ij} - \overline{Y} = \mu_i + \epsilon_{ij} - \overline{Y}$$

and $E(_2\epsilon_{ij}) = \mu_i - \mu$   where $\mu = \sum_i n_i \mu_i / N$

(5.7)   from (5.6) plots of $_2\epsilon_{ij}$ against the $i^{th}$ treatment

$i = 1,2,...p$, will be random with mean $\mu_i - \mu$ (the
mean of each error for each treatment would be at
a different level)

<u>Suppose Model (1) is Correct</u> (but Model (2) is used)

(5.8) then by replacing $Y_{ij}$ with (1) yields

$$_2\epsilon_{ij} = Y_{ij} - \overline{Y} = \mu_i + \beta(X_{ij} - \overline{X}..) + \epsilon_{ij} - \overline{Y} \text{ and}$$

$E(_2\epsilon_{ij}) = \mu_i - \mu + \beta(X_{ij} - \overline{X}..)$   where $\mu = \sum_i n_i \mu_i / N$

(5.9)   from (5.8) the plot for each $i^{th}$ treatment $i = 1,2,..p$,
of $_2\epsilon_{ij}$ against $X_{ij} - \overline{X}..$ will be random about the line
with slope $\beta$ and intercept $\mu_i - \mu$

6

<u>Conclusion</u>:

The plots of the error terms which are simply the scattergrams of Y scaled through $\overline{Y}$ gives the analyst an inclination as to the appropriate model. The MSE which is the estimated variance of the response variable determines the degree of improvement which may be had by using alternative models. Obviously if all the error plots are random and the MSE is sufficiently small, then either this simplistic model is adequate or other explanatory variables should be considered.

6. <u>Null Hypothesis: Model (3)</u> $(Y_{ij} = \alpha + \beta X_{ij} + \epsilon_{ij})$

<u>Estimation</u>

(6.1) estimate $\alpha$ and $\beta$ by the usual least squares estimates $\hat{\alpha}_3$ and $\hat{\beta}_3$

(6.2) compute the prediction errors $_3\epsilon_{ij} = (Y_{ij} - \hat{\alpha}_3 - \hat{\beta}X_{ij})$ which are the residuals from the regression using least squares

(6.3) compute $MSE_3 = \sum_j \sum_i (Y_{ij} - \hat{\alpha}_3 - \hat{\beta}_3 X_{ij})^2 / N{-}2$

(6.4) $_3\epsilon_{ij}$ is a new variable representing the responses corrected for the linear effect of X <u>assuming no treatment effect</u>

(6.5) the MSE is the estimated variance of the new variable <u>assuming no treatment effect</u>

(6.6) $\alpha_3$ and $\beta_3$ are unbiased estimates <u>assuming no treatment effect</u>

<u>Suppose Model (1) is Correct</u> (but Model (3) is used)

(6.7) then $Y_{ij} - \mu_i = \beta(X_{ij} - \overline{X}..) + \epsilon_{ij}$ is the true linear relationship, i.e., the responses corrected for their treatment means are linear in X

7

(6.8) based on (6.7), $\hat{\alpha}_3$ and $\hat{\beta}_3$ are biased estimates of $\alpha$ and $\beta$

(6.9) by replacing $Y_{ij}$ with (1) yields

$$_3\varepsilon_{ij} = Y_{ij} - \hat{\alpha}_3 - \hat{\beta}_3 X_{ij}$$

$$= \mu_i + \beta(X_{ij} - \overline{X}..) + \varepsilon_{ij} - \hat{\alpha}_3 - \hat{\beta}_3 X_{ij}$$

$$= \mu_i + \beta(X_{ij} - \overline{X}..) - \overline{Y} - \hat{\beta}_3(X_{ij} - \overline{X}..) + \varepsilon_{ij}$$

$$= (\mu_i - \overline{Y}) + (\beta - \hat{\beta}_3)(X_{ij} - \overline{X}..) + \varepsilon_{ij} \text{ and}$$

$$E(_3\varepsilon_{ij}) = \mu_i - \mu + \gamma (X_{ij} - \overline{X}..) \text{ where}$$

$$\gamma = \beta - E(\hat{\beta}_3) = \text{bias of } \hat{\beta}_3$$

(6.10) from (6.9) plots of $_3\varepsilon_{ij}$ within each treatment against $X_{ij}$ will be random about a line with slope $\gamma$ , the bias of $\hat{\beta}_3$

### Conclusion:

Model (3) is the traditional linear regression model. The least squares estimate for $\beta$ is biased when there is a underlying treatment effect which is not considered by the model. The extent of this bias can be determined by interpreting the graph from 6.10.

7. Null Hypothesis: Model (4) $(Y_{ij} = \mu_i + \varepsilon_{ij})$

### Estimation

(7.1) estimate $\mu_i$ by $\overline{Y}_{i.} = \sum\limits_{j=1}^{n_i} Y_{ij}/n_i$

(7.2) compute the prediction errors $_4\varepsilon_{ij} = Y_{ij} - \overline{Y}_{i.}$

(7.3) compute $MSE_4 = \sum\limits_i \sum\limits_j (Y_{ij} - \overline{Y}_{i.})^2/N-p$

(7.4) $MSE_4$ is the estimated variance of the response variable after it has been corrected for the mean treatment effect assuming no covariate effect.

8

<u>Suppose Model (1) is Correct</u>   (but Model (4) is used)

(7.5)   by replacing $Y_{ij}$ with (1) yields

$$_4\varepsilon_{ij} = Y_{ij} - \overline{Y}_{i.}$$

$$= \mu_i + \beta(X_{ij} - \overline{X}_{..}) + \varepsilon_{ij} - \overline{Y}_{i.}$$

but $\overline{Y}_{i.} = \mu_i + \beta(\overline{X}_{i.} - \overline{X}_{..}) + \overline{\varepsilon}_{i.}$

hence $_4\varepsilon_{ij} = \beta(X_{ij} - \overline{X}_{i.}) + \varepsilon_{ij} - \overline{\varepsilon}_{i.}$

and $E(_4\varepsilon_{ij}) = \beta(X_{ij} - \overline{X}_{i.})$

(7.6)   from (7.5) the plot of $_4\varepsilon_{ij}$ against $(X_{ij} - \overline{X}_{i.})$ will be
        random about the line   $\beta(X_{ij} - \overline{X}_{i.})$

## Conclusion:

Model (4) is the usual one way ANOVA.  The inadequacies of this
model, i.e., the need to add a covariate variate can be determined by
interpreting graph (7.6).  The $_4\varepsilon_{ij}$ term is the new response variable corrected
for the treatment effect without considering the covariate effect.

8.  <u>Null Hypothesis: Model (1)</u> $(Y_{ij} = \mu_i + \beta(X_{ij} - \overline{X}_{..}) + \varepsilon_{ij})$

## Estimation

(8.1)   unlike the previous models discussed, Model (1) considers
        both the covariate and treatment effect concurrently.

9

(8.2) the estimates of the parameters $\mu_1$ and $\beta$ are simultaneously developed in order to conform to comment (8.1)

(8.3) Estimate $\mu_1$ in terms of $\beta$

$$Y_{ij} = \mu_1 + \beta(X_{ij} - \overline{X}..) + \epsilon_{ij}$$

$$\overline{Y}_{i\cdot} = \mu_1 + \beta(\overline{X}_{i\cdot} - \overline{X}..) + \overline{\epsilon}_{i\cdot}$$

$$\mu_1 = \overline{Y}_{i\cdot} - \beta(\overline{X}_{i\cdot} - \overline{X}..) + \overline{\epsilon}_{i\cdot}$$

and an unbiased estimator of $\mu_1$ would be

$$_1\widehat{\mu}_1' = \overline{Y}_{i\cdot} - \beta(\overline{X}_{i\cdot} - \overline{X}..)$$

(8.4) Estimate $\beta$ in terms of $\mu_1$

$$Y_{ij} = \mu_1 + \beta(X_{ij} - \overline{X}..) + \epsilon_{ij}$$

$$Y_{ij} - \mu_1 = \beta(X_{ij} - \overline{X}..) + \epsilon_{ij}$$

and hence an unbiased estimator for $\beta$ would be the least squares estimate $\widehat{\beta}_1'$ reflecting the linear relationship of the response variable corrected for the treatment effect.

(8.5) Since $\mu_1$ and $\beta$ are unknown, simultaneous estimates are derived using the relationships expressed in 8.3 and 8.4. Replacing $\mu_1$ with $_1\widehat{\mu}_1'$ in 8.4 yields

$$Y_{ij} - _1\widehat{\mu}_1' = \beta(X_{ij} - \overline{X}..) + \epsilon_{ij}$$

10

$$Y_{ij} - \overline{Y}_{i\cdot} = \beta(\overline{X}_{i\cdot} - \overline{X}..) - \beta(X_{ij} - \overline{X}..) + \varepsilon_{ij} \text{ and}$$

$$Y_{ij} - \overline{Y}_{i\cdot} = \beta(X_{ij} - \overline{X}_{i\cdot}) + \varepsilon_{ij}$$

which is the estimated relationship expressed in (8.4)

And $\widehat{\beta}_1$ the least squares estimate of the above linear relationship
is the sought after estimate of $\beta$ which takes into consideration the treatment effect.

And $_1\widehat{\mu}_i = \overline{Y}_{i\cdot} - \widehat{\beta}_1(\overline{X}_{i\cdot} - \overline{X}..)$ is the sought after estimate of $\mu_i$
which takes into consideration the covariate effect

(8.6)  Compute the prediction error

$$_1\varepsilon_{ij} = Y_{ij} - [_1\widehat{\mu}_1 + \widehat{\beta}_1(X_{ij} - \overline{X}..)]$$

$$= Y_{ij} - \overline{Y}_{i\cdot} + \widehat{\beta}_1(\overline{X}_{i\cdot} - \overline{X}..) - \widehat{\beta}_1(X_{ij} - \overline{X}..)$$

$$= (Y_{ij} - \overline{Y}_{i\cdot}) - \widehat{\beta}_1(X_{ij} - \overline{X}_{i\cdot}) \text{ which}$$

are the residuals from regression of the transformed
variables in (8.5)

Compute $MSE_1 = \sum\limits_{j=1}^{n_i} \sum\limits_{i=1}^{q} [(Y_{ij} - \overline{Y}_{i\cdot}) - \beta_1(X_{ij} - \overline{X}_{i\cdot})]^2/N-p-1$

which is the MS of the residuals from (8.5).

(8.7)  The plots of $_1\varepsilon_{ij}$ against $X_{ij}$ and against the various
treatments illustrate the adequacy of the model.

## Conclusions:

As the general form for all the previous models:
Comparisons of $MSE_1$ with $MSE_2$, $MSE_3$, $MSE_4$ would indicate the
improvement of model 1 over models 2,3,4 respectively.

11

Graphs of the error terms also would indicate the improvement of model 1 over models 2,3,4. An estimate of the treatment effect can be had by comparing $\overline{Y}_{i.} - \hat{\beta}_1 (\overline{X}_{i.} - \overline{X}..)$ for every i.

Note: The analysis presented here is for investigative purposes only. As with most graphical techniques, the investigator should be aware of the possible deformities caused by an arbitrary choice of scales.

## 9. Traditional Treatment of Analysis of Covariance

The traditional analysis of covariance applies the analysis of variance to the data after it has been corrected for the covariate or regression effect. Relating this method to what has been previously described yields the following relationships.

The total $SS_t = (MSE_3)(N-2)$ is the sum of squares of the response variable corrected for the regression effect under the assumption or null hypothesis of no treatment effect.

The within $SS_w = (MSE_1)(N-p-1)$ is the pooled within treatment sum of squares of the response variable corrected for the regression effect. In this case both the covariate and treatments are simultaneously considered.

The between SS $= SS_t - SS_w$ is the sum of square of the treatment mean of the response variable corrected for the regression effect.

12

The corresponding ANOVA Table is as follows:

## ANOVA TABLE

| Source of Variation | Degrees of Freedom | Sum of Squares | Mean Squares |
|---|---|---|---|
| Between Groups | $p-1$ | $SS_B = SS_t - SS_w$ | $SS_B/p-1$ |
| Within Groups | $\left(\sum_{i=1}^{p} n_i\right)-p-1$ | $SS_w$ | $MSE_1$ |
| Total | $\left(\sum_{i=1}^{p} n_i\right)-2$ | $SS_t$ | $MSE_3$ |

The F test, $(SS_t - SS_w)/p-1/SS_w/\left(\sum_{i=1}^{p} n_i\right)-p-1$, rejects the null hypothesis of no treatment effect if $MSE_1$ is sufficiently smaller then $MSE_3$, i.e., the addition of the treatment effect to model (3) yielding Model (1) sufficiently improves the model.

Calculations for the ANOVA Table from any standard regression package can be done as follows:

(9.1)  Regress $Y_{ij}$ on $X_{ij}$ using linear model.

(9.2)  SS residuals = $SS_t$ in the ANOVA

(9.3)  Regress $(Y_{ij} - \overline{Y}_{i.})$ on $(X_{ij} - \overline{X}_{i.})$ using linear model

(9.4)  SS residuals = $SS_w$ in the ANOVA

(9.5)  Test significance of $\beta$ from 9.3
      if significant then compute ANOVA Table as above
      if insignificant then there is no covariant effect , and the
      traditional one way ANOVA table should be used.

13

## 10. Example

To further explain this technique, a graphical example will be used to illustrate the various stages and interpretations of the analysis. Figure 1 represents the Y responses to two treatments. (The underlying model used to contrive this example contains both treatment and covariate effects but for simplicity does not contain the error terms. Hence a correct model should perfectly fit the data with zero error).

Applying comments 5.5, 5.7 and 5.9 to figure (1)[*] indicates that model (1) may be the appropriate model, i.e., within each treatment, the errors are linear with the same slope.

Figure (2) illustrates the plot of the residuals $_3\varepsilon_{1j}$ from the least squares calculations assuming model 3. (The least squares line is plotted against the scatter in figure 1.)

Applying comments 6.10, to figure (2) indicates that $\hat{\beta}_3$ assuming no treatment effect is biased with $\gamma = -.5$ (The estimate $\hat{\beta}_3 = 2.506$).
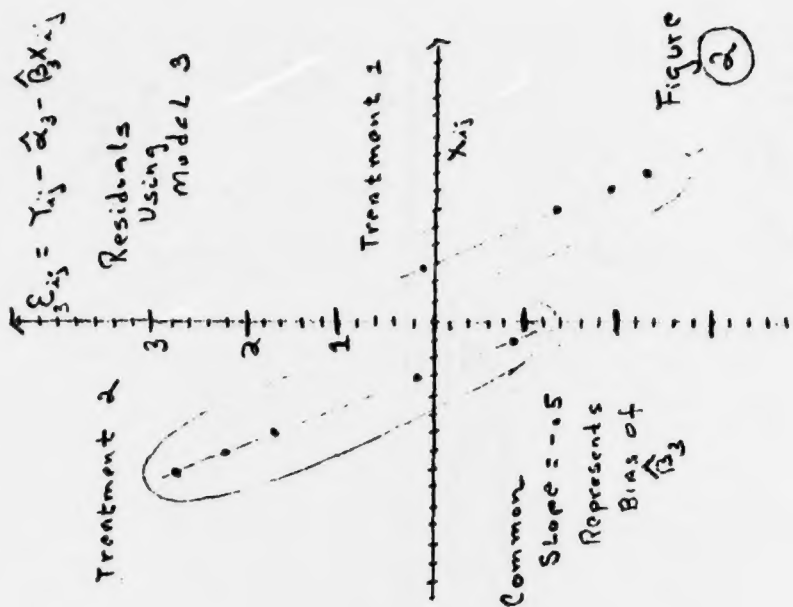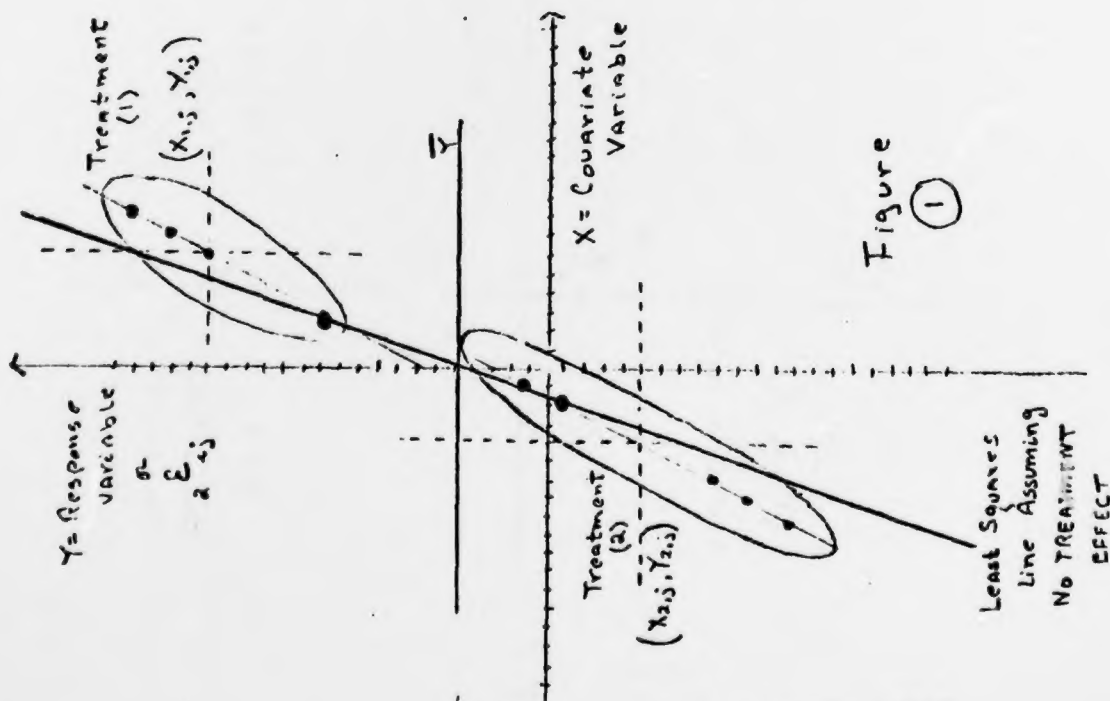
Figure (3) illustrates the plot of $_4\varepsilon_{1j} = (Y_{1j} - \overline{Y}_{1.})$ against $(X_{1j} - \overline{X}_{1.})$ which assumes model (4) and corrects the data for the treatment effect.
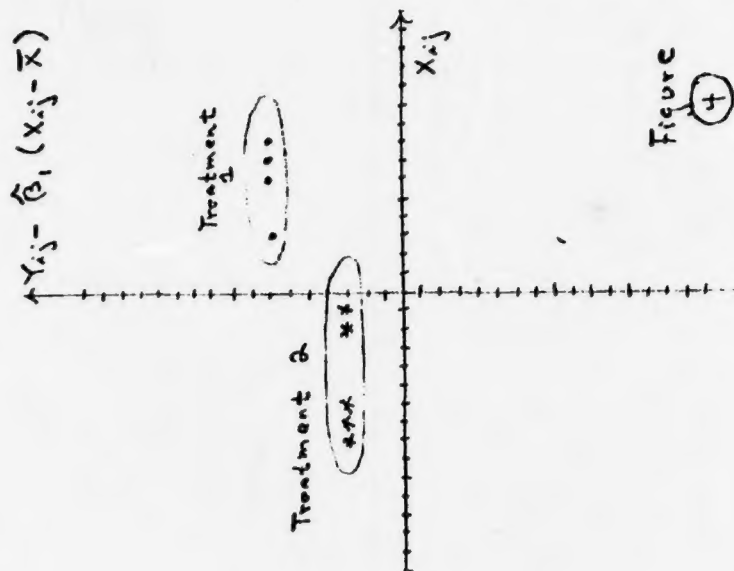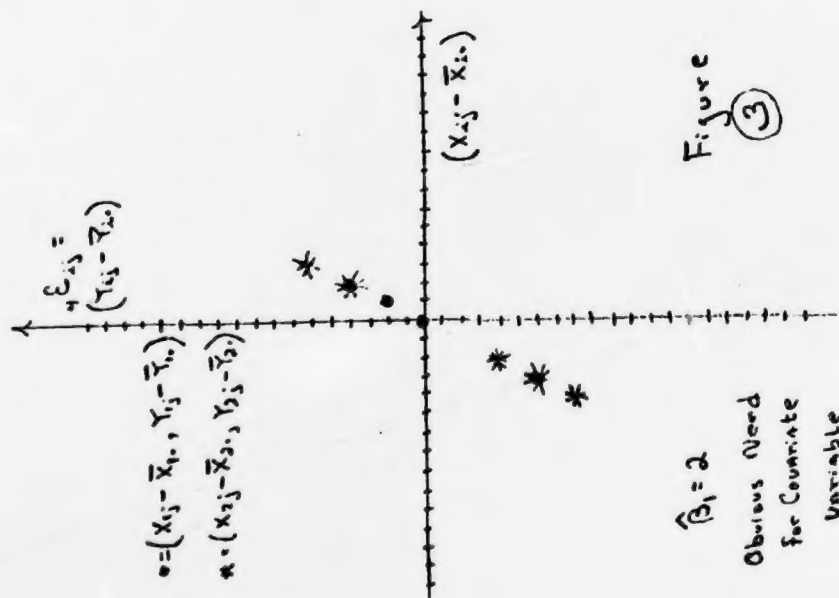
Applying comment 7.6, there is an obvious need for a covariate variable X and $\hat{\beta}_1 = 2$.

Figure (4) is equivalent to figure(2)except that the estimate $\hat{\beta}_1$ was used assuming treatment effect. The plot $Y_{1j} - \hat{\beta}_1(X_{1j} - \overline{X}..)$ against $X_{1j}$ represents the response corrected for the covariate effect and shows the actual treatment effect.

Other figures such as those noted in 8.7 should also be considered, but in this case would simply be a constant zero since the error term was omitted from the example.

---

[*] Since the response variable $Y_{1j}$ is a linear translation of $_2\varepsilon_{1j}$ the comments 5.5, 5.7, and 5.9 are applicable to this figure.

Figure (6)

$\mathcal{E}_{ij} = Y_{ij} - \hat{\alpha}_3 - \hat{\beta}_3 x_{ij}$

Residuals Using Model 3

Treatment 1

Treatment 2

Common Slope = -.5 Represents Bias of $\hat{\beta}_3$



Figure (1)

Treatment (1) $(x_{1j}, Y_{1j})$

Y = Response Variable or $\mathcal{E}_{ij}$

X = Covariate Variable

Treatment (2) $(x_{2j}, Y_{2j})$

Least Squares Line Assuming No Treatment Effect

$Y_{ij} - \hat{\beta}_1(X_{ij} - \bar{X})$

$X_{ij}$

Treatment 1

Treatment 2

Figure 1

$\hat{\varepsilon}_{ij} = (Y_{ij} - \bar{Y}_{i\cdot})$

$(X_{ij} - \bar{X}_{i\cdot})$

$\bullet = (X_{1j} - \bar{X}_{1\cdot}, Y_{1j} - \bar{Y}_{1\cdot})$

$* = (X_{2j} - \bar{X}_{2\cdot}, Y_{2j} - \bar{Y}_{2\cdot})$

$\hat{\beta}_1 = 2$

Obvious Need for Covariate Variable

Figure 2

16

## 11. <u>Conclusions</u>:

The graphical technique described in this paper is not only a method of model selection but is also an illustrative description of the rationals underlying Analysis of Covariance. Extensions of this methodology can easily be made to cover situations where multiple treatment and/or co-variate variables are encountered.

## REFERENCES

1 Dixon, Wilfred V. and Frank J. Massey, "Introduction to Statistical Analysis," McGraw-Hill, New York, 1957.

2 Gotwals, Edwin P., "Use of End Item Age and Usage in Demand Forecasting," DRC Inventory Research Office, December 1976.

3 Huitson, Alan, "The Analysis of Variance," Hafner Publishing, New York, 1966.

4 Johnson, J., "Econometric Methods," McGraw-Hill, New York, 1972.

5 Scheffe, Henry, "The Analysis of Variance," John Wiley and Sons, New York, 1959.

6 Snedecor, George W. and William G. Cochran, "Statistical Methods," Iowa State University Press, Ames, Iowa, 1972.